

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Provinsi Jawa Barat tahun 2018. Data yang diperoleh dari BKKBN adalah data tentang *stunting* beserta faktor-faktor yang mempengaruhi *stunting* pada balita di enam Kabupaten/Kota di Jawa Barat yaitu Kabupaten Bandung, Kota Bandung, Kabupaten Garut, Kabupaten Bandung Barat, Kabupaten Subang, dan Kabupaten Sumedang.

3.2 Analisis Data

Dalam analisis kluster pada tahap pengklasteran dibedakan menurut tipe data yang dimiliki. Pada umumnya analisis kluster terfokus pada data numerik, namun tidak memungkiri bahwa ada pula kasus dengan data kategorik bahkan terdapat kasus dengan campuran data kategorik dan numerik. Analisis kluster pada data kategorik tidak dapat diperlakukan seperti data numerik. Hal tersebut dikarenakan sifat khusus dari data kategorik sendiri, sehingga pengklasteran data kategorik menjadi lebih rumit dibandingkan dengan pengklasteran untuk data numerik (Hair, Black, Babin, & Anderson, 2010).

3.2.1 Pengklasteran Data Kategorik

Pengklasteran pada data kategorik dilakukan dengan menggunakan ukuran keserupaan atau jarak. Dalam data kategorik pengklasteran dapat dilakukan dengan metode hirarki maupun non hirarki. Pengklasteran menggunakan metode hirarki dan non hirarki dinilai tidak tepat digunakan untuk data kategorik sehingga dikembangkan metode *ROCK* yang digunakan untuk pengklasteran data kategorik tersebut (Guha, Rastogi, & Shim, 2000).

Awal mula metode *ROCK* dikembangkan dari metode pengklasteran hirarki yang digunakan untuk data kategorik. Metode *ROCK* pertama kali dikemukakan oleh Guha, Rastogi dan Shin pada tahun 2000. Karena pada pengklasteran hirarki digunakan jarak antara titik untuk pengklasteran dinilai

kurang sesuai untuk data kategorik, maka dibentuk sebuah kosep baru yakni *link* yang digunakan untuk mengukur kesamaan antara sepasang titik data. Kelebihan dari metode ROCK adalah dapat menangani adanya pencilan dengan cukup efektif. Pemangkasan pencilan memungkinkan untuk membuang yang tidak ada tetangga, sehingga titik tersebut tidak berpartisipasi dalam pengklasteran. Namun dalam beberapa situasi pencilan dapat hadir sebagai kelompok-kelompok kecil (Guha, Rastogi, & Shim, 2000).

Pengklasteran data kategorik menggunakan algoritma *ROCK* dilakukan dengan tiga langkah. Langkah pertama adalah menghitung *similaritas* menggunakan rumus koefisien *Jaccard* (Rahayu, 2013). Ukuran keserupaan antara pasangan objek ke- i dan objek ke- j dihitung dengan rumusan yang didefinisikan pada persamaan 3.1.

$$sim(X_i, X_j) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|}, \quad i \neq j \quad (3.1)$$

dengan,

$i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, n$

X_i : Himpunan pengamatan ke- i dengan $X_i = \{x_{1i}, x_{2i}, x_{3i}, \dots, x_{m_{kategorik^i}}\}$,

X_j : Himpunan pengamatan ke- j dengan $X_j = \{x_{1j}, x_{2j}, x_{3j}, \dots, x_{m_{kategorik^j}}\}$,

$|X|$: Bilangan kardinal atau jumlah anggota dari himpunan X ,

Langkah kedua adalah menentukan tetangga, pengamatan dinyatakan sebagai tetangga jika nilai $sim(X_i, X_j) \geq \theta$. Langkah ketiga adalah menghitung *Link* antar objek pengamatan. Besarnya *Link* dipengaruhi oleh nilai *threshold* (θ) yang merupakan parameter yang ditentukan oleh pengguna yang dapat digunakan untuk mengontrol seberapa dekat hubungan anatar objek. Besarnya nilai θ yang dimasukkan adalah $0 < \theta < 1$. Ketika tidak ada lagi *link* antara kelompok-kelompok atau jumlah dari kelompok yang diharapkan sudah terpenuhi maka algoritma *ROCK* akan berhenti. (Dutta, Mahanta, & Arun, 2005).

Metode *ROCK* menggunakan informasi tentang *link* sebagai ukuran kemiripan antar objek. Jika terdapat objek pengamatan X_i , X_j dan X_k di mana X_i merupakan tetangga dari X_j , dan X_j merupakan tetangga X_k , maka dinyatakan X_i memiliki *link* dengan X_k walaupun X_i bukan tetangga dari X_k . Cara untuk menghitung *link* untuk semua kemungkinan pasangan dari n objek dapat menggunakan matriks **A**. Matriks **A** merupakan matriks yang berukuran $n \times n$ yang bernilai 1 jika X_i dan X_j dinyatakan serupa (tetangga) dan bernilai 0 dan jika X_i dan X_j tidak mirip (bukan tetangga). Jumlah *link* antar pasangan X_i dan X_j diperoleh dari hasil kali antara baris ke X_i dan kolom ke X_j dari matriks **A**. Jika *link* antara X_i dan X_j semakin besar maka semakin besar pula kemungkinan X_i dan X_j berada dalam satu klaster yang sama (Dutta, Mahanta, & Arun, 2005).

Selanjutnya untuk penggabungan klaster dengan menggunakan algoritma *ROCK* didasarkan atas ukuran kebaikan (*goodness measure*) antar klaster dengan rumus pada persamaan 3.2. *Goodness measure* adalah persamaan yang menghitung jumlah *link* dibagi dengan kemungkinan *link* yang terbentuk berdasarkan ukuran klasternya (Tyagi & Sharma, 2012).

$$g(C_i, C_j) = \frac{\text{link}[C_i, C_j]}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}} \quad (3.2)$$

di mana dengan $\text{link}[C_i, C_j] = \sum_{X_i \in C_i, X_j \in C_j} \text{link}(X_i, X_j)$ yang menyatakan jumlah *link* dari semua kemungkinan pasangan objek yang ada dalam C_i dan C_j dimana C_i dan C_j adalah klaster ke- i dan j , serta n_i dan n_j masing-masing menyatakan jumlah anggota dalam kelompok ke- i dan j , sedangkan $f(\theta) = \frac{1-\theta}{1+\theta}$.

3.2.2 Pengklasteran Data Numerik

Pengklasteran untuk data numerik dilakukan berdasarkan ukuran ketidakserupaan atau jarak untuk data numerik. Hasil pengklasteran biasanya disajikan ke dalam bentuk diagram pohon (*dendogram*) yang memungkinkan penelusuran objek-objek yang diamati menjadi informatif dan lebih mudah.

Refki Ardy Prasetya, 2020

ANALISIS KLASSTER DENGAN METODE *ENSEMBLE ROCK* UNTUK DATA CAMPURAN

(Studi Kasus *Stunting* di Provinsi Jawa Barat)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Metode yang akan digunakan dalam pengklasteran untuk data numerik adalah hirarki.

Terdapat dua teknik pengelompokan dalam analisis kluster hirarki yakni secara agglomeratif (teknik penggabungan) dan *divisive* (teknik pembagian). Metode pengklasteran secara *divisive* dimulai dengan mengansumsikan bahwa seluruh objek berada dalam satu kluster selanjutnya objek dipecah ke dalam kluster yang lebih kecil lagi sehingga terbentuk kluster sebanyak jumlah keseluruhan objek yang diklasterkan. Sedangkan untuk metode pengklasteran secara agglomeratif merupakan kebalikan dari pengklasteran metode secara *divisive*, yakni dimulai dengan mengansumsikan bahwa setiap objek dikelompokkan ke dalam kelompok yang lebih besar sehingga semua objek berada di dalam satu kelompok.

Pada umumnya proses pengklasteran yang digunakan adalah pengklasteran secara agglomeratif, langkah awal pengklasteran secara agglomeratif adalah menentukan dua objek yang memiliki jarak yang paling dekat kemudian menjadikan objek-objek tersebut ke dalam satu kluster. Langkah kedua adalah menghitung kembali jarak antar kluster kemudian menggabungkan kluster yang memiliki jarak terdekat. Langkah ketiga dan selanjutnya adalah sama dengan langkah kedua, yakni menghitung ulang jarak antar kluster kemudian menyatukan kluster yang memiliki jarak yang paling dekat. Langkah terakhir adalah menggabungkan dua buah kluster besar yang telah terbentuk menjadi satu kluster besar.

Perbedaan pengklasteran jika menggunakan metode agglomeratif satu dengan yang lainnya adalah pada langkah kedua, yaitu di mana keputusan mengenai kluster mana yang akan digabungkan. Keputusan tersebut bergantung dari metode hirarki mana yang akan digunakan. Pada umumnya ada 5 metode hirarki yang sering digunakan (Sharma, 1996) yaitu :

1. Metode *Single Linkage* atau *Nearest-Neighbor*
2. Metode *Complete Linkage* atau *Farthest-Neighbor*
3. Metode *Average Linkage*
4. Metode *Ward* atau *Minimum-Variance*

Refki Ardy Prasetya, 2020

**ANALISIS KLASSTER DENGAN METODE *ENSEMBLE ROCK* UNTUK DATA CAMPURAN
(Studi Kasus *Stunting* di Provinsi Jawa Barat)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

5. Metode *Centroid Linkage*

Algoritma pengklasteran di atas akan dipaparkan sebagai berikut :

a. Metode *Single Linkage* atau *Nearest-Neighbor*

Metode *Single Linkage* adalah proses pengklasteran yang didasarkan pada jarak terdekat antar objeknya. Jika dua objek terpisah oleh jarak yang terkecil atau minimum, maka kedua objek tersebut akan bergabung menjadi satu kluster dan seterusnya (Rachmatin, 2014). Adapun ukuran jarak yang digunakan adalah :

$$d_{w(u,v)} = \min(d_{wu}, d_{wv}) \quad (3.3)$$

Hasil keseluruhan dari algoritma *Single Linkage* dapat disajikan dalam bentuk *dendogram* , dengan cabang mengartikan kluster (Johnson & Whincern, 2007).

b. Metode *Complete Linkage* atau *Farthest-Neighbor*

Metode *Complete Linkage* adalah metode pengklasteran yang didasarkan pada jarak terjauh antar objek. Jika dua objek terpisah oleh jarak yang jauh, maka kedua objek tersebut akan digabung menjadi satu kluster hingga seterusnya (Rachmatin, 2014). Adapun ukuran jarak yang digunakan adalah :

$$d_{w(u,v)} = \max(d_{wu}, d_{wv}) \quad (3.4)$$

c. Metode *Average Linkage*

Metode *Average Linkage* adalah metode pengklasteran yang didasarkan pada jarak rata-rata antar objeknya. Dengan n_u dan n_v merupakan jumlah pengamatan dalam kluster ke- u dan ke- v (Rachmatin, 2014). Adapun ukuran jarak yang digunakan adalah :

$$d_{w(u,v)} = \frac{n_u}{n_u+n_v} d_{wu} + \frac{n_v}{n_u+n_v} d_{wv} \quad (3.5)$$

d. Metode *Ward* atau *Minimum-Variance*

Metode *Ward* atau *Minimum-Variance* bertujuan untuk memperoleh kelompok yang memiliki varians internal kluster yang sekecil mungkin. Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis varians untuk menghitung jarak antar kluster atau metode ini meminimumkan

jumlah kuadrat (ESS) (Rachmatin, 2014). Adapun ukuran jarak yang digunakan adalah :

$$ESS = \sum_{j=1}^n x_j^2 - \frac{1}{n} (\sum_{j=1}^n x_j)^2 \quad (3.6)$$

e. Metode *Centroid Linkage*

Metode *Centroid Linkage* merupakan metode pengklasteran dengan memperhatikan rata-rata dari setiap objek yang bergabung berdasarkan jarak minimum yang diperoleh dari matriks jarak *Euclid* (Rachmatin, 2014).

Adapun ukuran jarak yang digunakan adalah :

$$D(M, N) = d(\overline{y_M}, \overline{y_N}) \quad (3.7)$$

Di mana, $\overline{y_M}, \overline{y_N}$ merupakan centroid dari M dan N, $\overline{y_M}, \overline{y_N}$ dapat dicari menggunakan rumus berikut :

$$\overline{y_M} = \sum_{i=1}^{n_M} \frac{y_i}{n_M} \quad (3.8)$$

$$\overline{y_N} = \sum_{i=1}^{n_N} \frac{y_i}{n_N} \quad (3.9)$$

(Rencher, 2002).

Sebenarnya metode apapun yang akan digunakan pada metode hirarki bertujuan untuk meminimalkan variansi dalam kluster dan memaksimalkan variansi antar tiap kluster (Simamora, 2005).

3.2.3 Pengklasteran Data Campuran

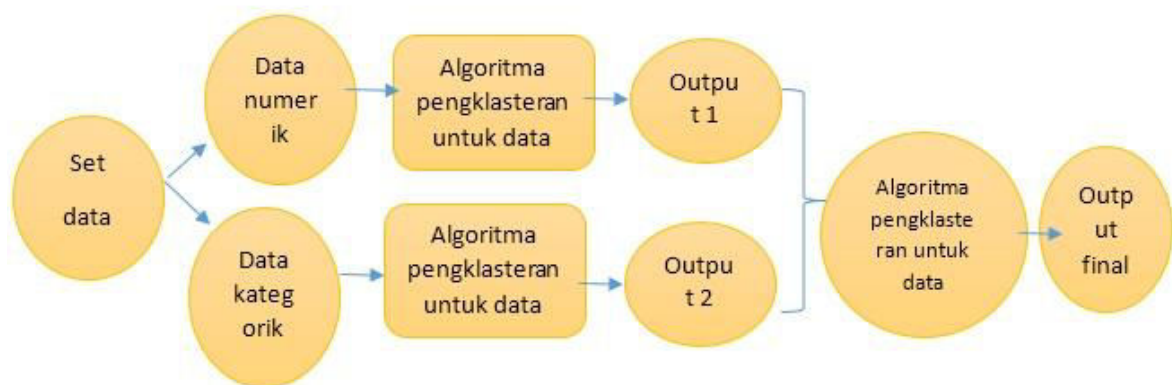
Dalam pengklasteran dalam data bertipe campuran, akan digunakan tiga macam proses algoritma pengklasteran yaitu algoritma pengklasteran untuk menganalisis data numerik, algoritma pengklasteran untuk menganalisis data kategorik, dan yang terakhir algoritma pengklasteran untuk menganalisis data numerik dan kategorik (campuran).

Salah satu algoritma yang bisa diterapkan pada data yang bertipe campuran adalah algoritma *Cluster Ensemble Based Mixed Data Clustering (CEBMDC)* yang merupakan suatu algoritma pengklasteran dengan pendekatan kluster *ensemble*. Kluster *ensemble* adalah suatu metode yang digunakan untuk

menggabungkan dan menjalankan beberapa algoritma pengklasteran yang berbeda untuk mendapatkan partisi yang sama dari data tersebut, serta bertujuan untuk menyatukan hasil dari hasil-hasil pengklasteran masing-masing (Hee, Xu, & Deng, 2005)

Menurut He, Xu & Deng (2005) mengungkapkan bahwa hasil dari masing-masing algoritma pengklasteran dapat digabungkan menjadi data yang baru dengan tipe data kategorik. Karena hasil output dari masing-masing algoritma pengklasteran merupakan data bertipe kategorik, sehingga masalah kluster *ensemble* dapat dipandang sebagai masalah dari pengklasteran data kategorik.

Adapun skema pengklasteran ensemble menggunakan algoritma *CEBMDC* sebagai berikut :



Gambar 3.1. Langkah dari Algoritma CEBMDC

Langkah-langkah yang dilakukan dalam analisis data campuran menggunakan metode pengklasteran ensemble yang disebut algoritma *CEBMDC* memiliki prosedur sebagai berikut , (Hee, Xu, & Deng, 2005).

1. Pisahkan data menjadi dua subdata, yaitu data numerik dan data kategorik.
2. Lakukan pengklasteran terhadap data numerik menggunakan algoritma pengklasteran untuk data numerik yaitu metode hirarki agglomeratif.
3. Lakukan pengklasteran terhadap data kategorik menggunakan algoritma pengklasteran untuk data kategorik yaitu metode ROCK.

4. Gabungkan output pengklasteran dari masing masing algoritma dan dianggap data kategorik, proses ini disebut proses *ensemble*.
5. Lakukan pengklasteran *ensemble* menggunakan ROCK terhadap data baru yang dianggap data kategorik untuk mendapat klaster akhir.

3.2.4 Evaluasi Hasil Pengklasteran

Setelah didapat hasil dari pengklasteran selanjutnya penilaian kualitas struktur data dari klaster yang terbentuk menggunakan jenis ukuran validitas. Untuk mencari tahu klaster optimal dalam analisis klaster dengan data bertipe numerik adalah dengan ukuran *index Dunn* yang diajukan oleh *J.C.Dunn*. Klaster yang terpisah pada umumnya memiliki jarak antara klaster yang besar dan jarak dalam klaster kecil (Satato, Khotimah, & Muhammad, 2015). *Indeks Dunn* tidak memiliki suatu rentang nilai, untuk mencari nilai *indeks Dunn* yang terbaik atau optimum dapat dilihat dari nilai terbesar yang dihasilkan (Dewanti, 2013) begitupun menurut Bolshaka & Azuaje (2001) nilai terbesar dari *indeks Dunn* diambil sebagai jumlah optimum klaster. Berikut adalah rumus *indeks Dunn* :

$$D(c) = m_{1 \leq i \leq n} \left\{ m_{1 \leq i \leq n, i \neq j} \left\{ \frac{d(c_i, c_j)}{m_{1 \leq k \leq n} (d'(c_k))} \right\} \right\} \quad (3.10)$$

Di mana :

$d(c_i, c_j)$ = jarak antara c_i dan c_j

$d'(c_k)$ = jarak dalam klaster c_k

Salah satu pengukuran validitas metode yang terbaik dapat dengan melihat nilai rasio S_b dan S_w . Dengan menggunakan nilai rata-rata dari variabel , simpangan baku antar kelompok (S_b) dan simpangan baku di dalam kelompok (S_w) untuk data numerik sebagai berikut :

$$S_w = \frac{1}{C} \sum_{c=1}^C S_c \quad (3.11)$$

dengan

C : jumlah klaster yang terbentuk.

Refki Ardy Prasetya, 2020

**ANALISIS KLASSTER DENGAN METODE *ENSEMBLE ROCK* UNTUK DATA CAMPURAN
(Studi Kasus *Stunting* di Provinsi Jawa Barat)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

S_c : simpangan baku kelompok ke- c .

$$S_b = \left[\frac{1}{C-1} \sum_{c=1}^C (\bar{x}_c - \bar{x})^2 \right]^{\frac{1}{2}} \quad (3.12)$$

dengan

C : jumlah klaster yang terbentuk.

\bar{x}_c : rata-rata kelompok ke- c .

\bar{x} : rata – rata keseluruhan kelompok.

Semakin kecil dari nilai rasio S_b dan S_w , maka klaster yang dihaikan semakin baik.

Alvionita (2017) mengungkapkan Pengukuran validitas untuk data kategorik dapat menggunakan tabel kontingensi yang ekuivalen dengan melakukan ANOVA (*Analysis of Variance*), ANOVA yang dimaksud adalah ANOVA khusus untuk pengujian keragaman data kategorik dalam ruang lingkup analisis klaster. Untuk data kategorik ukuran keragaman telah dikembangkan oleh beberapa ahli diantaranya Light dan Margolin (1971), Okada (1999), dan Kader serta Perry (2007). Jika terdapat sebanyak n pengamatan dengan n_k merupakan jumlah pengamatan daengan kategori ke- k di mana $k = 1, 2, 3, \dots, K$ dan $\sum_{k=1}^K = n$. Selanjutnya n_{kc} merupakan jumlah pengamatan dengan kategori ke- k dan kelompok ke- c , di mana $c = 1, 2, 3, \dots, C$ dengan C adalah jumlah kelompok yang terbentuk , sehingga $n_c = \sum_{k=1}^K n_{kc}$ merupakan jumlah pengamatan pada kelompok ke- c dan $n_k = \sum_{c=1}^C n_{kc}$ merupakan jumlah pengamatan pada kategori ke- k , sehingga total jumlah pengamatan dapat dituliskan menjadi

$$n = \sum_{c=1}^C n_c = \sum_{k=1}^K n_k = \sum_{k=1}^K \sum_{c=1}^C n_{kc}$$

Jumlah kuadrat (SST) untuk sebuah variabel dengan data kategorik dapat dirumuskan sebagai berikut :

$$SST = \frac{n}{2} - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (3.13)$$

Sedangkan untuk jumlah total kuadrat dalam kelompok (SSW) dapat dirumuskan sebagai berikut :

Refki Ardy Prasetya, 2020

**ANALISIS KLASSTER DENGAN METODE *ENSEMBLE ROCK* UNTUK DATA CAMPURAN
(Studi Kasus *Stunting* di Provinsi Jawa Barat)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

$$SSW = \sum_{c=1}^C \left(\frac{n_c}{2} - \frac{1}{2n_c} \sum_{k=1}^K n_{kc}^2 \right) = \frac{n}{2} - \frac{1}{2} \sum_{c=1}^C \frac{1}{n_c} \sum_{k=1}^K n_{kc}^2 \quad (3.14)$$

Dan jumlah kudrat antar kelompok (SSB) dapat dirumuskan sebagai berikut :

$$SSB = \frac{1}{2} \left(\sum_{c=1}^C \frac{1}{n_c} \sum_{k=1}^K n_{kc}^2 \right) - \frac{1}{2n} \sum_{k=1}^K n_k^2 \quad (3.15)$$

Mean of squares total (MST), *mean of squares within (MSW)* , dan *mean of squares between (MSB)* dapat dirumuskan pada seperti:

$$MST = \frac{SST}{(n - 1)} \quad (3.16)$$

$$MSW = \frac{SSW}{(n - C)} \quad (3.17)$$

$$MSB = \frac{SSB}{(C - 1)} \quad (3.18)$$

Serta simpangan baku dalam kelompok (S_w) dan simpangan baku dalam kelompok (S_b) untuk data kategorik dapat dirumuskan seperti :

$$S_w = [MSW]^{\frac{1}{2}} \quad (3.19)$$

$$S_b = [MSB]^{\frac{1}{2}} \quad (3.20)$$

Sama halnya dengan data numerik, jika semakin kecil dari nilai rasio S_b dan S_w , maka klaster yang dihaikan semakin baik.

3.3 Teknik Analisis Data

Pada penelitian ini digunakan algoritma CEBMDC dengan langkah-langkah prosedur sebagai berikut (Guha, Rastogi, & Shim, 2000):

1. Pisahkan data, antara data kategorik dan numerik.
2. Pengklasteran untuk peubah kategorik menggunakan metode *ROCK*
 - a) Lakukan inisialisasi objek sebagai klaster dengan anggota tunggal
 - b) Bentuk keserupaan antar objek dengan kriteria menggunakan persamaan 3.1

- c) Tentukan *threshold* (θ). Nilai (θ) yang digunakan dalam penelitian ini yaitu $\theta = 0,05$, $\theta = 0,10$, $\theta = 0,15$, $\theta = 0,20$, $\theta = 0,30$, $\theta = 0,5$, $\theta = 0,6$, dan $\theta = 0,75$.
 - d) Hitung nilai *link* antar objek pengamatan
 - e) Hitung nilai *goodness measure* menggunakan persamaan 3.2 sehingga didapat klaster yang diinginkan.
 - f) Ulangi langkah (e) dengan nilai θ yang berbeda.
 - g) Hitung nilai rasio S_w dan S_b untuk masing-masing nilai θ dengan rumus pada persamaan 3.19 dan 3.20.
 - h) Bandingkan masing-masing nilai rasio S_w dan S_b dan menentukan pengklasteran terbaik yakni dengan melihat nilai rasio S_w dan S_b yang terkecil.
 - i) Tentukan keanggotaan klaster berdasarkan pengelompokan terbaik.
3. Pengklasteran untuk peubah numerik menggunakan metode hirarki agglomerative.
- a) Lakukan inisialisasi objek sebagai klaster dengan anggota tunggal
 - b) Tentukan ukuran keserupaan dengan jarak *euclid* dengan rumus pada persamaan 2.1.
 - c) Gabungkan klaster yang memiliki jarak terdekat.
 - d) Perbaharui matriks jarak terdekat dengan menggunakan metode *single linkage* menggunakan persamaan 3.3
 - e) Ulangi langkah (c) dan (d) sampai hanya terbentuk 1 klaster
 - f) Ulangi langkah (a) dan (e) menggunakan metode *complete linkage* seperti pada persamaan 3.4
 - g) Ulangi langkah (a) dan (e) menggunakan metode *average linkage* seperti pada persamaan 3.5
 - h) Ulangi langkah (a) dan (e) menggunakan metode *ward* seperti pada persamaan 3.6

- i) Ulangi langkah (a) dan (e) menggunakan metode *centroid* seperti pada persamaan 3.7
- j) Hitung indeks validitas kelompok menggunakan *indeks Dunn* seperti pada persamaan 3.10
- k) Hitung nilai rasio S_w dan S_b dari pengelompokan yang telah diperoleh dari e, f, g, h dan i.
- l) Bandingkan hasil langkah (j) dan menentukan pengelompokan terbaik yaitu pengelompokan dengan nilai rasio S_w dan S_b terkecil.
- m) Tentukan keanggotaan klaster berdasarkan pengklasteran terbaik.

4. Gabungan hasil klaster (tahapan *ensemble*)

Setelah mendapatkan klaster yang terbaik dari masing-masing metode, tahapan selanjutnya adalah melakukan penggabungan klaster. Tahapan ini sama dengan melakukan klasterisasi pada data katgorik menggunakan metode *ROCK* dengan nilai *threshold* (θ) yang digunakan dalam penelitian ini $\theta = 0,05$, $\theta = 0,10$, $\theta = 0,15$, $\theta = 0,20$, $\theta = 0,30$, $\theta = 0,5$, $\theta = 0,6$, dan $\theta = 0,75$. Di mana input untuk tahapan ini adalah klaster pengelompokan data numerik dengan metode hirarki agglomeratif (output 1) dan klaster dari pengelompokan data kategorik denga metode *ROCK* (output 2). Output 1 dan output 2 dinyatakan sebagai peubah kategorik yang digunakan untuk menyusuk final klaster. Final klaster yang baik adalah jumlah klaster yang memiliki nilai rasio S_w dan S_b terkecil rumus pada persamaan 3.19 dan 3.20.